

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Sociology Department, Faculty Publications

Sociology, Department of

2020

How Do Interviewers and Respondents Navigate Sexual Identity Questions in a CATI Survey?

Jerry Timbrook

University of Nebraska-Lincoln, jtimbrook2@huskers.unl.edu

Jolene Smyth

University of Nebraska-Lincoln, jsmyth2@unl.edu

Kristen M. Olson

University of Nebraska-Lincoln, kolson5@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/sociologyfacpub>



Part of the [Family, Life Course, and Society Commons](#), and the [Social Psychology and Interaction Commons](#)

Timbrook, Jerry; Smyth, Jolene; and Olson, Kristen M., "How Do Interviewers and Respondents Navigate Sexual Identity Questions in a CATI Survey?" (2020). *Sociology Department, Faculty Publications*. 743.
<https://digitalcommons.unl.edu/sociologyfacpub/743>

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Sociology Department, Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

How Do Interviewers and Respondents Navigate Sexual Identity Questions in a CATI Survey?

Jerry Timbrook, Jolene D. Smyth, and Kristen Olson

University of Nebraska-Lincoln, Lincoln, NE, USA

Corresponding author — J. Timbrook;
emails jtimbrook2@huskers.unl.edu; jsmyth2@unl.edu; kolson5@unl.edu

Background

Survey-based research has demonstrated that sexual minority individuals experience unique outcomes in areas such as physical and mental health (Boehmer et al. 2007; Hatzenbuehler 2014, 2017), crime (Herek 2009), public education (Kosciw et al. 2015), same-sex romantic relationships and family (Powell and Downey 1997; Umberson et al. 2015), and economics (Black et al. 2007). Having a reliable and valid measure of sexual identity (i.e., the way in which an individual self-describes their sexual orientation) (Gagnon and Simon 1973) is essential for conducting research on sexual minorities. Indeed, many national surveys such as the General Social Survey, the National Health Interview Survey, and the National Survey of Family Growth ask survey respondents about their sexual identity.

Published as Chapter 10 in Philip S. Brenner (ed.), *Understanding Survey Methodology: Sociological Theory and Applications*, Frontiers in Sociology and Social Research 4, (2020), pp 219-245.

doi:10.1007/978-3-030-47256-6_10

Copyright © Springer Nature Switzerland AG 2020. Used by permission

The percentage of US adults identifying as a sexual minority has increased from 3.5% (8.3 million) in December 2012 to 4.1% (10.05 million) in December 2016 (Gates 2017). As the prevalence of sexual minorities is still low, inaccurate answers or item nonresponse to sexual identity questions (SIQs) may result in large distortions of estimates of sexual minorities. In a meta-analysis, Ridolfo et al. (2012) demonstrate that item nonresponse rates for SIQs range from 1.6% to 4.3%, with an average of approximately 2%. For context, this places item nonresponse for SIQs higher than education questions (1.1%), but below income questions (11.2%) (Conron et al. 2008). The threat of item nonresponse due to concerns over question sensitivity has led many researchers to advocate against the use of interviewers to administer SIQs (SMART 2009; Ridolfo et al. 2012). Instead, they recommend that survey researchers ask SIQs using only self-administered modes of data collection (i.e., using mail surveys, web surveys, or Computer-Assisted Self-Interview [CASI] devices for face-to-face surveys) (SMART 2009; Ridolfo et al. 2012).

Yet asking SIQs exclusively in a self-administered context may not be a feasible approach. First, self-administered modes are not without drawbacks of their own: mail surveys are time consuming, face-to-face surveys using CASI are time consuming and expensive, and web surveys do not have a sampling frame with adequate coverage of the US population (Dillman et al. 2014). In contrast, researchers using telephone surveys can collect nationally representative data quickly. Second, while SIQs provide important demographic information, they are often not the primary focus of a survey. Many private survey companies (e.g., Pew, Gallup, Abt Associates) and government surveys (e.g., the Behavioral Risk Factor Surveillance System) rely on telephone surveys to achieve a variety of cost, quality, and timeliness objectives; these organizations are not likely to switch modes to improve data quality for a single demographic question like sexual identity. Thus, understanding the implications of administering SIQs in all modes, including telephone, is of continued interest.

Previous research on SIQs in telephone surveys has been limited to examinations of data quality indicators such as “Don’t Know” or refusal answers (e.g., VanKim et al. 2010; Ridolfo et al. 2012; Fredriksen-Goldsen and Kim 2015). These studies do not consider interactional problems between the interviewer and the respondent that may occur prior

Do you consider yourself to be heterosexual or straight, gay or lesbian, or bisexual?

- 1 HETEROSEXUAL OR STRAIGHT
- 2 GAY OR LESBIAN
- 3 BISEXUAL
- 4 OTHER (SPECIFY)
- 8 DK
- 9 REF

Fig. 1 Sexual Identity Question

to negotiating an answer. Yet investigating when and why the interaction between interviewers and respondents deviates from a paradigmatic “question asked/question answered” sequence can provide researchers with an in-depth understanding of problems with questions in interviewer-administered surveys (Schaeffer and Maynard 1996).

In this study, we explore interactional difficulties that result from telephone administration of an SIQ to gain insight into the problems associated with this question. Specifically, we focus on two potential sources of strain: (1) the sensitivity of the question and (2) the construction of the question. For example, the stigma associated with having a sexual minority status may lead some respondents to provide intentionally inaccurate answers or refuse to answer the question entirely (Sylva et al. 2010; Stange 2014). Issues with the question’s construction may similarly lead some respondents to unintentionally misclassify their sexual identity (e.g., if they are unfamiliar with words like “heterosexual”) (Miller and Ryan 2011; Ridolfo et al. 2011). Additionally, telephone surveys are a cognitively taxing aural medium (de Leeuw 2005; Dillman et al. 2014) in which respondents may experience problems with complex questions, requiring extra help to complete these items.

The construction of the SIQ we use in this study (**Fig. 1**) was recommended by the Sexual Minority Assessment Research Team (SMART 2009). This SIQ contains wording that is similar to many SIQs used in federal surveys.¹ Using this SIQ, we examine the following research questions:

¹ For an overview of these measures, see the Federal Interagency Working Group on Improving Measurement of SOGI 2016 report.

RQ1: Do respondents express concerns with the sensitivity of the sexual identity question? In what ways do they express sensitivity concerns?

RQ2: Do respondents express concerns with the construction of the sexual identity question? In what ways do they express question construction concerns?

RQ3: How do interviewers react to concerns over sensitivity and question construction?

RQ4: What respondent and interviewer characteristics are associated with problems with the sensitivity and construction of the sexual identity question?

Below, we consider how the sensitivity and the construction of the SIQ may differentially affect three classes of actors involved in administering an SIQ: (1) lesbian, gay, and bisexual (LGB) respondents,² (2) heterosexual respondents, and (3) interviewers.

Question Sensitivity

LGB respondents. LGB respondents may experience *stigma* when being asked about their sexual identity. Stigma is a sense of shame associated with an identity, condition, or status that is viewed as undesirable by society (Goffman 1963). More specifically, LGB respondents may fear experiencing sexual stigma if they associate themselves with a non-heterosexual identity (Herek 2009, 2011). As the social presence of another individual can activate concerns over stigma, interviewer-administered telephone surveys may trigger an LGB respondent's fear of being stigmatized (Goffman 1963; de Leeuw 2005).

Nearly a quarter of US adults believe that homosexuality should be discouraged (Pew Research Center 2017), making it reasonable for sexual minority respondents to expect some degree of disapproval if they disclose their identity to an interviewer. Fear of sexual stigmatization

² We recognize that not all sexual minorities have a lesbian, gay, or bisexual identity. However, for the sake of brevity, we use the phrase "LGB respondents" to refer to all respondents with a minority sexual identity. We also do not examine transgender respondents or responses to questions about gender identity.

may be signaled when LGB respondents express reluctance before selecting “gay or lesbian” or “bisexual” as their sexual identity. Reluctance may take the form of explicit comments about discomfort with or the intrusiveness of the SIQ (e.g., “That’s a personal question.”), or paralinguistic expressions of discomfort such as stuttering, disfluencies, or laughing (e.g., Lavin and Maynard 2001, 2002).

Instead of answering the SIQ honestly, LGB respondents may instead choose to conceal³ their identity, potentially to avoid discrimination that may result from disclosing a stigmatized sexual identity (Goffman 1959, 1963; Sylva et al. 2010; Stange 2014). This may take the form of intentional misclassification, where LGB respondents select “heterosexual or straight” (the non-stigmatized identity) as their answer to the SIQ. Alternatively, LGB respondents may conceal their LGB identity by refusing to answer the SIQ altogether.

Finally, LGB respondents may be *unaffected* by the sensitivity of the SIQ. Assuming that they do not have issues with the construction of the SIQ, these respondents would answer the question paradigmatically using the “gay or lesbian” or “bisexual” response options.

Heterosexual respondents. Unlike LGB respondents, heterosexual respondents are not likely to consider their sexual identity stigmatized, as it reflects the majority (Goffman 1963; Herek 2009, 2011). However, these respondents may still consider the SIQ to be *intrusive* as it asks about sexual or romantic attraction to others, a topic that many consider private (Tourangeau et al. 2000). Similar to LGB respondents, heterosexual respondents may signal their discomfort with the question’s sensitivity through explicit comments about the private nature of the question or with paralinguistic cues. These respondents may ultimately answer the question truthfully, or they may conceal their sexual identity by refusing to answer the question due to its sensitive nature (Tourangeau and Yan 2007).

Heterosexual respondents who are uncomfortable with the concept of homosexuality or bisexuality may express *backlash* when faced with an LGB-inclusive context (Hooten et al. 2009; Stange 2014; Stange et al. 2018). This backlash may take the form of explicit anti-LGB comments when asked to indicate their sexual orientation. For example,

³ We note that identity concealment cannot be observed in an observational dataset such as ours. We mention this here to describe what *may* happen when an SIQ is administered.

religious beliefs play an important role in the formation of individuals' opinions about homosexuality (Olson et al. 2006; Adamczyk and Pitt 2009), with major religious texts providing the basis for societal disapproval of homosexuality (Yip 2005). Thus, backlash may be expressed as religious-based disapproval of homosexuality (e.g., "The bible says that's wrong.").

Respondents may also express backlash to the SIQ if they fear being associated with the stigmatized LGB identity. This experience, known as homophobia (Anderson 2009), may cause heterosexual respondents to dis-identify from being LGB rather than selecting the appropriate "heterosexual or straight" response option. These respondents may instead dis-identify from an LGB identity by providing answers such as "I'm not gay," "I'm normal," or "I'm 100% straight."

Some respondents who express backlash, despite their disapproval, may ultimately provide an accurate and acceptable answer. Other backlash respondents may refuse to answer the question in protest. Finally, it is also possible that in light of increasing rates of acceptance of homosexuality (Pew Research Center 2017), heterosexual respondents will have no issue with the sensitivity of the SIQ. These *unaffected* respondents will provide honest, acceptable answers to the question with no additional commentary.

Interviewers. An essential part of an interviewer's job is to ensure that respondents provide acceptable answers to each question in a survey, and continue to participate (Fowler and Mangione 1990; Schaeffer et al. 2010). Interviewers who demonstrate responsiveness to respondents' concerns during the survey interview cultivate a sense of rapport that can motivate respondents to provide accurate and thoughtful answers (Cannell et al. 1981; Dijkstra 1987; Garbarski et al. 2016). However, sensitive questions like the SIQ may represent a threat to rapport with the potential to derail respondents' continued and optimal participation in a survey.

To contend with the intrusive nature of sensitive questions, interviewers may engage in a variety of *rapport-building* behaviors to foster continued interaction with their respondents (Japac 2008; Garbarski et al. 2016). First, interviewers may attempt to minimize disruption to rapport by distancing themselves from the researcher. This can manifest through explicit comments such as "they make me ask these questions" (Houtkoop-Steenstra 2000) or when interviewers

apologize to respondents for the topic (e.g., “I’m so sorry about this question”) (Garbarski et al. 2016). Second, interviewers may seek to minimize the time spent on the SIQ. In such cases, interviewers may tell respondents they do not have to answer the question (e.g., offering the Don’t know/Refusal response options) (Japiec 2008). Similarly, interviewers may choose not to probe unacceptable answers to the SIQ (e.g., “I’m not gay”) in an effort to move on to the next question as quickly as possible. These interviewers may incorrectly (but intentionally) code unacceptable answers as actual answers to avoid negotiating an acceptable response (Ridolfo et al. 2012). Finally, interviewers may skip asking the SIQ entirely to avoid an uncomfortable interaction, and falsify the respondent’s answer to the question.

Interviewer behaviors change as they gain more experience with a particular survey (van der Zouwen et al. 1991; Olson and Peytchev 2007; Olson and Bilgen 2011). When administering the SIQ, interviewers may find that certain rapport-building strategies are more useful than others, making within-survey experience of interest. It is also possible that interviewers with more years on the job have more experience administering SIQs. Thus, these interviewers may have already cultivated strategies for coping with the sensitive and complex nature of the SIQ.

We also note that interviewers themselves are a component of the survey context. Thus, different characteristics of the interviewer may be perceived as more or less accepting of an LGB sexual identity. Accordingly, these characteristics may make LGB respondents more or less likely to accurately disclose their identity, or may increase or decrease the amount of backlash expressed by heterosexual respondents. For example, women are generally more supportive of homosexuality than men (Herek 2002). Thus, LGB respondents may be more likely to disclose their identity to female interviewers, while backlash respondents may feel more comfortable expressing their anti-LGB opinions to male interviewers. In addition, although we cannot assess it with the data from our study, because men are more likely to report anti-LGB attitudes to interviewers they perceive as being heterosexual (Kemph and Kasser 1996) backlash may be more common from male respondents when they believe their interviewer is heterosexual. Conversely, LGB respondents may be more comfortable discussing issues related to their sexual minority status with LGB interviewers

(Rotheram-Borus et al. 1994) and may therefore be more apt to report being a sexual minority to interviewers they perceive as being LGB.

Construction of the SIQ

LGB respondents. The construct of sexual identity is more complex than straight, gay, lesbian, and bisexual. Many individuals possess sexual identities that do not easily conform to categorization, identities that may be fluid depending on time and place (Gamson 1995; Diamond 2008). For example, Ridolfo et al. (2012) describe a transitioning transgender respondent who preferred the sexual identity of “queer” to describe their simultaneous status as a homosexual man and a heterosexual woman. Respondents with these minority sexual identities may find that the response options to the SIQ are *inapplicable* to them (Ridolfo et al. 2012).

Respondents who do not self-identify as possessing any of the sexual identities expressed by the SIQ’s response options may instead choose to endorse “Other” (Ridolfo et al. 2012; Eliason et al. 2016). However, in telephone surveys, this response option is not always offered as a part of the question stem; these respondents must reveal that their sexual identity does not conform to any of the response options provided, and the interviewer would then offer the “Other” response option. Thus, some respondents may intentionally misclassify their sexual identity or skip the question to avoid further interaction.

Heterosexual respondents. Fears of stigmatization stemming from having a minority status means that sexual identity is often salient for LGB individuals. However, individuals who occupy majority positions across social identities (e.g., gender, race, sexuality, class) rarely consider their identities due to their assumed normality (e.g., Perry 2001). Thus, heterosexual respondents without a salient sexual identity may find the SIQ’s response options to be unfamiliar (Miller and Ryan 2011; Ridolfo et al. 2012).

Confusion over definitions in the question may prevent some respondents from providing an accurate answer, thereby compromising data integrity (Fowler 1995; Tourangeau et al. 2000). For example, cognitive interviews of SIQs have demonstrated that the words “gay,” “lesbian,” and “straight” are well understood by English-speaking respondents, but the word “heterosexual” is not (Miller and Ryan

2011). This may lead some respondents to express confusion with the SIQ's response options (Ridolfo et al. 2012). These respondents may unintentionally misclassify themselves as having a sexual identity of "Other," or skip the SIQ altogether (Ridolfo et al. 2012). Miller and Ryan (2011) also find that some heterosexual respondents may misunderstand the "bisexual" response option, believing that it expresses "one man and one woman." In the same study, some respondents conflated their sexual and gender identities (e.g., answering "I'm female"). Respondents who are older and less educated may be most likely to be unfamiliar with these terms (Ridolfo et al. 2012).

Respondents for whom sexual identity is not salient may, instead of endorsing an unfamiliar response option, answer the SIQ by dis-identifying from an LGB identity (Miller and Ryan 2011). In these cases, heterosexual respondents may indicate that they are "normal" or "not gay." Thus, dis-identification may not exclusively indicate distancing oneself from LGB due to homophobia (as discussed above). Instead, disidentification may also characterize respondents who wish to indicate that they do not have a minority sexual identity, but also do not understand the response options.

Words commonly used in SIQs may not have an equivalent meaning across different languages and cultures. For example, Spanish-speaking respondents in cognitive interviews have reported difficulty understanding translations of the words "gay," "lesbian," "bisexual," "homosexual," and "heterosexual" (Ridolfo and Schoua-Glusberg 2009; Miller and Ryan 2011; Michaels et al. 2017). Additionally, the vocabulary that members of the LGB community use to identify their sexual minority status differs across cultures (Kulick 2000). For instance, the word "gay" in Japanese conflates same-sex attraction and gender identity, referring to men who dress in women's clothing *and* are attracted to other men (Valentine 1997; McLelland 2000). Misunderstandings due to cultural/linguistic differences may lead respondents to unintentionally misclassify their sexual identity, or skip the SIQ (Miller and Ryan 2011; Ridolfo et al. 2012; Michaels et al. 2017). Although the present study uses only interviews conducted in English, this body of literature highlights the importance of pretesting translations of SIQs before administering them across cultures and languages.

Aural administration of the SIQ. Questions with complex syntax (e.g., questions with multiple clauses) can impair respondents'

understanding of a question (Tourangeau et al. 2000). The SIQ in this study contains three main response options (clauses): (1) heterosexual or straight, (2) gay or lesbian, or (3) bisexual. Two of these clauses are further subdivided by the conjunction “or” (i.e., “heterosexual or straight” and “gay or lesbian”). This complex syntax may result in confusion that negatively impacts the accuracy of data collected using the SIQ (Ridolfo et al. 2011). Additionally, issues of question construction can disproportionately affect respondents with lower working memory capacity, such as those with lower educational attainment (Ceci 1991) and older adults (Salthouse 1991). Thus, the SIQ may be particularly taxing to respondents in these subgroups.

Asking questions in the aural channel of telephone surveys can magnify issues of complex syntax (de Leeuw 2005; Dillman et al. 2014). For example, respondents in a telephone survey may hear the clause “heterosexual or straight” and assume that heterosexual and straight are different response options as they are separated by the conjunction “or” (Ridolfo et al. 2011). Respondents may then think they are subsequently being asked if they are gay “or” lesbian, which can add additional confusion. This may lead some respondents to answer the SIQ before the full stem has been read by the interviewer (Dillman et al. 2014). For example, upon hearing “Do you consider yourself heterosexual or straight,” respondents may prematurely answer with either “heterosexual” or “straight”, or “yes” or “no”.

Interviewers. Interviewers are often required to read questions exactly as worded in order to reduce variable interviewer effects on data (Fowler and Mangione 1990). However, interviewers sometimes make changes to questions that they feel are difficult for respondents to understand (Japac 2008; Dykema et al. 2020). Thus, changes when reading the SIQ to respondents may indicate an interviewer’s attempt to repair the complexity of the question. These changes can be minor (i.e., changes that do not affect question meaning), or they can be major (i.e., changes that do affect question meaning). As an example, interviewers may add numbers to response options in the question stem to delimit them for respondents (e.g., “Do you consider yourself to be *one* heterosexual or straight, *two* gay or lesbian, or *three* bisexual?”) (Haan et al. 2013). Changes in question reading may be more prevalent for experienced interviewers who are less likely to read questions exactly as worded (Fowler and Mangione 1990).

Measuring Problems with Sensitivity and Question Construction of the SIQ

Previous research has investigated construction issues with the SIQ in both lab and field studies. Miller and Ryan (2011) and Ridolfo et al. (2009, 2011, 2012) use cognitive interviewing, a qualitative lab-based technique (Willis 2005), to explore how lesbian, gay, bisexual, transgender, and heterosexual respondents interpret SIQs. To ensure representation, these studies oversampled minority sexual respondents.

Although these studies provide rich detail on how respondents interpret the question, cognitive interviews do not replicate the questioning asking and answering experience that occurs in a field study. Additionally, sexual minority respondents were generally recruited from LGBT community centers, likely communicating an LGB-inclusive context to respondents. This also means that LGB respondents who may be more concerned with sexual stigma were likely excluded from these studies. Similarly, these studies used purposive recruitment and were sponsored by the federal government. Thus, these studies may have attracted heterosexual respondents who were more accepting of LGB identities or were generally unlikely to express backlash over the SIQs.

Ridolfo et al. (2012) and Fredriksen-Goldsen and Kim (2015) explore respondent reactions to SIQs in field studies by examining patterns of response (i.e., item nonresponse, and selecting “Other” response options). These results come from a naturalistic interviewer/respondent interaction, overcoming a criticism of cognitive interviews. In their analysis, the authors focus mainly on issues of question construction that lead to these patterns of response. As respondents are likely to skip questions that are confusing, researchers often use rates of item nonresponse to proxy for respondent difficulty with question construction (Krosnick 1991). However, respondents may also skip a question they find sensitive (Tourangeau and Yan 2007). Thus, investigating patterns of response alone does not enable researchers to disentangle issues of question construction and sensitivity.

In this study, we attempt to remedy the lack of representativeness in cognitive interviews and lack of in-depth understanding from the item nonresponse analyses by using behavior coding to explore the interaction between respondents and interviewers during telephone

administration of the SIQ in a random sample national survey (Schaeffer and Maynard 1996). We identify when the interaction between interviewers and respondents deviates from a paradigmatic sequence, and classify the problem that caused the deviation (i.e., question sensitivity, construction, or both).

Data and Methods

Sample

The data for this study come from the Work and Leisure Today 2 (WLT2) survey, a dual-frame random-digit dial telephone survey of U.S. adults conducted by Abt SRBI during August and September of 2015. The survey consisted of 58 questions asking about respondents' use of leisure time, use of technology, and demographics. The survey had 902 respondents (AAPOR RR3 = 7.8%), but three cases were eliminated from analysis due to call quality issues. Further, in four cases, interviewers did not administer the SIQ to respondents. This leaves $n = 895$ cases for analysis.

After transcribing each interview, we extract all conversational turns⁴ related to the SIQ. Our unit of analysis is the full administration of this item (i.e., all conversational turns related to the asking and answering of this item). Notably, the SIQ was administered directly after respondents were asked about their marital/partner status. If they were married or partnered, respondents were also asked if that was to someone of the same sex or the opposite sex.

Behavior Codes

The majority of our measures for this study come from behavior coding the interaction between interviewers and respondents during administration of the SIQ. The behavior coding method is used to objectively and reliably identify when interviewers and respondents deviate from a paradigmatic “question asked/question answered” sequence

⁴ We define conversational turns as a period of uninterrupted speech by a single actor, with turns ending when an actor finished speaking or was interrupted by another actor. Instances of overlapping speech were counted as their own turns.

(Fowler and Cannell 1996; Schaeffer and Maynard 1996). This allows us to identify when an interaction problem occurred during administration of the SIQ and classify the problem as an indicator of difficulty with question sensitivity, construction, or both. With the exception of interviewer question asking behaviors, codes in our dataset are not mutually exclusive. Thus, each behavior code is assigned independently of all other codes.

Our first set of behavior codes proxy for respondent issues with the sensitivity of the SIQ (**Table 1**). For each case, we code whether or not (no = 0; yes = 1) the respondent ever refuses to answer the SIQ during the interaction, regardless of whether this was the final answer in the data set. As a measure of *intrusiveness*, we code whether or not the respondent ever makes an explicit comment that the SIQ asks about a sensitive/personal topic or is generally off-topic (e.g., that the question deviated from the survey topic of leisure activities). *Backlash* is represented by two codes: whether or not the respondent makes a comment about their religious beliefs in connection to homosexuality, or expresses discomfort or anger over homosexuality. We also create a summary variable of whether a respondent ever expresses any of these four indicators of sensitivity (=1) or not (=0).

Next, we code interviewer *rapport-building* behaviors that indicate reactions to the sensitive nature of the SIQ for each case. We code whether or not the interviewer ever (1) blames the researcher for having to administer the SIQ, (2) apologizes for having to ask the SIQ, (3) offers the Don't Know/Refusal (DK/REF) response options, or (4) neglects to probe an unacceptable answer (i.e., does not probe when they should have). Again we create a summary variable of whether an interviewer ever engages in any of these four sensitivity behaviors or not.

Our third set of codes proxy for respondent confusion with the SIQ's construction. First, we code whether or not (no = 0; yes = 1) respondents ever commented that they answered this question in a previous question (e.g., because the previous question asks about the gender of the respondent's marital partner). Next, we create variables for four respondent behaviors that indicate difficulty with the *definitions* of words in the question. We code whether or not respondents expressed confusion about (1) the difference between the words "heterosexual" and "straight", (2) the difference between sexual identity and sexual behavior (e.g., if a respondent indicates they are asexual

Table 1 Proportions, Kappas, and Percent Agreement for Each Behavioral Indicator

<i>Code</i>	<i>Mean</i>	<i>(SE)</i>	<i>Kappa</i>	<i>% Agreement</i>
<i>Indicators of question sensitivity</i>				
Respondent behaviors				
Refused to answer	1.8%	(0.4%)	1.00	100.0%
<i>Intrusiveness</i>				
Comments about question being sensitive/off topic	3.4%	(0.5%)	1.00	100.0%
<i>Backlash</i>				
Comments about religion	0.7%	(0.2%)	–	–
Comments about discomfort with homosexuality	1.2%	(0.4%)	–	–
Any respondent indicator of question sensitivity	5.1%	(0.6%)		
Interviewer behaviors				
<i>Rapport-building</i>				
Blames the researcher for the question topic	0.7%	(0.3%)	–	–
Apologizes for the question topic	0.3%	(0.2%)	–	–
Offers the DK/REF response options	0.4%	(0.2%)	1.00	100.0%
Should have probed, but did not	4.5%	(0.8%)	1.00	100.0%
Any interviewer indicator of question sensitivity	5.3%	(0.8%)		
<i>Indicators of question construction problems</i>				
Respondent behaviors				
Comments answer was given in a previous question	0.2%	(0.2%)	–	–
<i>Definition confusion</i>				
Heterosexual vs. straight	2.1%	(0.6%)	1.00	100.0%
Sexual identity vs. sexual behavior	0.6%	(0.4%)	1.00	100.0%
Sexual identity vs. gender	0.8%	(0.3%)	–	–
Confusion with any other word in the question	1.2%	(0.3%)	–	–
<i>Response option confusion</i>				
Positive answer (e.g., “yes”)	1.9%	(0.5%)	–	–
Negation (e.g., “no”)	4.2%	(0.7%)	0.66	98.8%
Any respondent indicator of construction problems	9.8%	(1.0%)		
Interviewer behaviors				
<i>Question asking</i>				
Asked exactly as worded	88.9%	(3.1%)	0.64a	82.7%
Asked with a minor change	7.2%	(1.5%)		
Asked with a major change	3.9%	(1.9%)		
Read numbers with response options	4.8%	(3.4%)	1.00	100.0%
Any interviewer indicator of construction problems	11.8%	(3.7%)		
Behaviors indicating question sensitivity or construction problems				
Comments with uncertainty of answer	2.9%	(0.5%)	–	–
<i>Dis-identification</i>				
Answers what is not (e.g., “I’m not gay”)	0.9%	(0.3%)	1.00	100.0%
Comments about degree of own sexuality	4.2%	(0.5%)	0.75	96.8%
Comments about heterosexual as normative	1.5%	(0.4%)	1.00	100.0%
<i>Respondent paralinguistic expressions</i>				
Stuttering	5.0%	(0.7%)	0.86	97.9%
Disfluencies	15.2%	(1.3%)	0.97	98.9%
Laughter	12.5%	(1.8%)	0.94	98.9%
<i>Interviewer paralinguistic expressions</i>				
Stuttering	3.5%	(0.7%)	0.86	97.9%
Disfluencies	12.7%	(2.8%)	1.00	100.0%
Laughter	9.1%	(2.2%)	1.00	100.0%
Any construction or sensitivity indicator by any actor	31.6%	(3.1%)		

a. One kappa value was created to assess the question asking codes as they are dependent upon one another

because they are too old to engage in sexual behavior), (3) the difference between sexual identity and gender (e.g., answering with one's gender), or (4) the definition of any other words in the question. Our final two codes for construction problems indicate whether or not the respondent confused the SIQ's *response options* for yes/no rather than multiple choice. We code for whether or not respondents give a positive answer (e.g., "yes") or a negation (e.g., "no" or "nope") as a final response. We conclude by creating a summary measure indicating whether or not any of the seven indicators of respondent construction confusion were present during the SIQ's administration.

We also code three variables indicating interviewer reactions to the confusing construction of the SIQ. If an interviewer's reading of the SIQ deviated from the questionnaire's wording, we coded whether or not that was a minor change (i.e., a change not affecting question meaning) or a major change (i.e., a change affecting question meaning). We also code whether or not interviewers added numbers to each response option when reading the question stem (e.g., "*one* heterosexual or straight, *two* gay or lesbian, or *three* bisexual."). Again we create an overall indicator of any interviewer problems with question construction.

Our final collection of behavior codes represents variables that could indicate problems with the sensitivity or the construction of the question. First, we code whether or not respondents made comments about the uncertainty of their answer. Next, we code for the absence or presence of three respondent indicators of dis-identification from an LGB identity: (1) answering the question in terms of what is *not* their sexual orientation (e.g., "I'm not gay"), (2) commenting that the degree of their sexuality is fully heterosexual (e.g., "I'm 100% straight"), or (3) referring to heterosexual or straight as being "normal". We also create variables for the three paralinguistic expressions for both respondents and interviewers. We code whether or not each actor ever stuttered, made a filler disfluency (e.g., "uh" or "um"), or laughed during administration of the SIQ. We conclude by creating an overall indicator of any problem with the SIQ's sensitivity or construction made by either the respondent or the interviewer. This indicator includes respondent uncertainty and dis-identification, but excludes paralinguistic expressions (as they may indicate behaviors other than problems with the SIQ's sensitivity/construction).

Inter-Coder Reliability

Two individuals with extensive experience coding interviewer-respondent interactions (the first author and a research assistant) coded the same 10% random subset of the full sample ($n = 94$ cases) on each variable. Kappa values and percent agreement for each variable were calculated in Stata version 14.2 using the `kap` command (third and fourth column of Table 1). Behaviors that have “–” instead of a kappa value did not occur in the 10% subsample, making it impossible for Stata to calculate kappa values. All variables had kappa values greater than .60. We note that three variables had kappa values $< .8$, but in these cases, this does not indicate low agreement. For example, the code indicating whether or not a respondent answered with a negation had a kappa of .66 (bordering moderate agreement) but had a percent agreement of 98.8%; the coders only disagreed on negation codes for 2 of 94 cases. Therefore, we also report percent agreement, an alternative to kappa for rare events, for each variable (Viera and Garrett 2005). Percent agreement exceeds 80% for each variable.

Respondent and Interviewer Characteristics

Our final set of variables are used to explore the association between respondent and interviewer characteristics (**Table 2**). Respondent measures come from the survey dataset itself. For our respondent characteristics, we code the respondent's: (1) age overall ($M = 54.1$) and by the categories 18–35 (18.7%), 36–55 (52.5%), and 66+ (28.8%); (2) level of education as high school graduate or less (31.2%), some college (26.2%), college graduate or higher (42.6%); and (3) gender as male (47.5%) or female (52.5%). Missing values to these variables were imputed to the mean/mode (percent missing $< 4\%$). We also code each respondent's sexual identity using their answer to the SIQ from the interview transcript (i.e., we do not use the answers recorded by interviewers). The majority of our sample was heterosexual or straight (90.6%), with 0.6% identifying as gay or lesbian, 2.4% identifying as bisexual, 0.1% identifying as “Other,” 1.8% providing DK or REF answers, and 4.5% of respondents providing uncodable answers (i.e., answers that did not fit one of the response options).

Table 2 Respondent and Interviewer Characteristics

<i>Respondent Characteristics</i>	
Number of respondents	895
Age	
18–35	18.7%
36–65	52.5%
66+	28.8%
Mean (in years)	54.1
Education	
High school graduate or less	31.2%
Some college	26.2%
College graduate or more	42.6%
Sex	
Male	47.5%
Female	52.5%
Sexual identity (coded from transcripts)	
Heterosexual or straight	90.6%
Gay or lesbian	0.6%
Bisexual	2.4%
Other	0.1%
Don't know/refused	1.8%
Uncodable answer	4.5%
<i>Interviewer characteristics</i>	
Number of interviewers	27
Interviewer tenure	
One year or less	70.4%
More than one year	29.6%
Sex	
Male	59.3%
Female	40.7%
Average # of R per Interviewer	34

Interviewer characteristics were obtained from the survey organization's administrative records. We code each of our $n = 27$ interviewers' tenure on the job as 1 year or less (70.4%) or more than 1 year (29.6%), and code each interviewer's gender as male (59.3%) or female (40.7%). Finally, we code interviewers' within-survey experience at the respondent level: each interview was coded as occurring within the first half of each interviewer's WLT2 cases, or the second half.

Analysis Methods

In telephone surveys, respondents are nested within interviewers (Hox 1994; Olson and Peytchev 2007; Olson and Bilgen 2011). Thus, the analyses in this study use the complex survey design procedures (svy procedures) in Stata 14.2 to account for the clustering of respondents within interviewers.

We start by computing proportions for each of our behavior coded variables of interest to explore the frequency with which each behavior occurs. Then, we compare our overall rates of problems with question sensitivity to overall rates of problems with question construction separately for respondents and interviewers using dependent tests of proportions.

Next, we use a logistic regression framework to examine the bivariate relationship between respondent's age and each of our four overall indicators (i.e., overall indicators of question sensitivity and construction problems for both respondents and interviewers). We analyze respondent age first as a categorical predictor (i.e., 18–35, 36–65, and 65+), and then as a continuous, mean-centered variable. We then use a logistic regression framework to examine the bivariate relationship between each of our four overall indicators and respondent education, interviewer tenure, interviewer sex, and interviewer within-survey experience. Finally, we estimate four multivariate logistic regression models, each one predicting whether or not one of our four overall indicators occurred. In these multivariate models, we include all respondent and interviewer characteristics as predictors. With one exception (footnoted in the Results section), none of our bivariate results change in these multivariate models.

Results

Indicators of Question Sensitivity

The first two columns of Table 1 present overall proportions of occurrence and standard errors for each behavior-coded variable in this study. The top path of **Fig. 2** displays a tree diagram of respondent behaviors indicating problems with the SIQ's sensitivity. First, 1.8% of respondents refused to answer the SIQ at some point during the

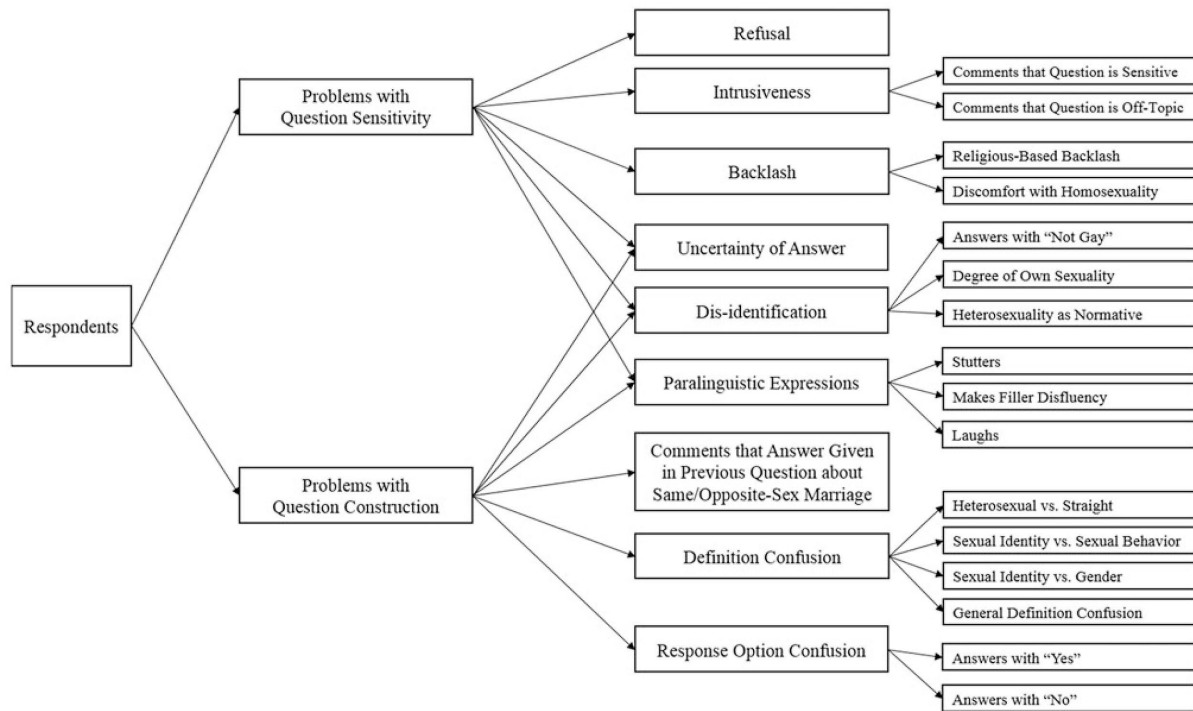


Fig. 2 Tree Diagram of Coded Respondent Behaviors

interaction, even if they ultimately provided an answer. For indicators that respondents perceived the SIQ to be *intrusive*, 3.4% of respondents commented that the question was sensitive (e.g., “Wow. We’re getting personal now huh?”) or off-topic (“Like, what does that have to do with leisure?”). The majority of these comments were made by heterosexual respondents (63.3% of the comments), however, 16.7% of these comments were made by respondents who ultimately refused to answer the question, and 16.7% were made by respondents who never provided an acceptable answer due to the interviewer’s failure to probe. Only 1 respondent who self-identified as bisexual (3.3% of the comments) commented that the question was intrusive, while no respondents self-identifying as gay or lesbian made such a comment. This indicates that respondents who identify as LGB generally have no trouble with the intrusiveness of the SIQ.

Only a small proportion of respondents expressed *backlash* to the SIQ. Few respondents (0.7%) made references to religion in connection with the SIQ, with some indicating that they are “what the good

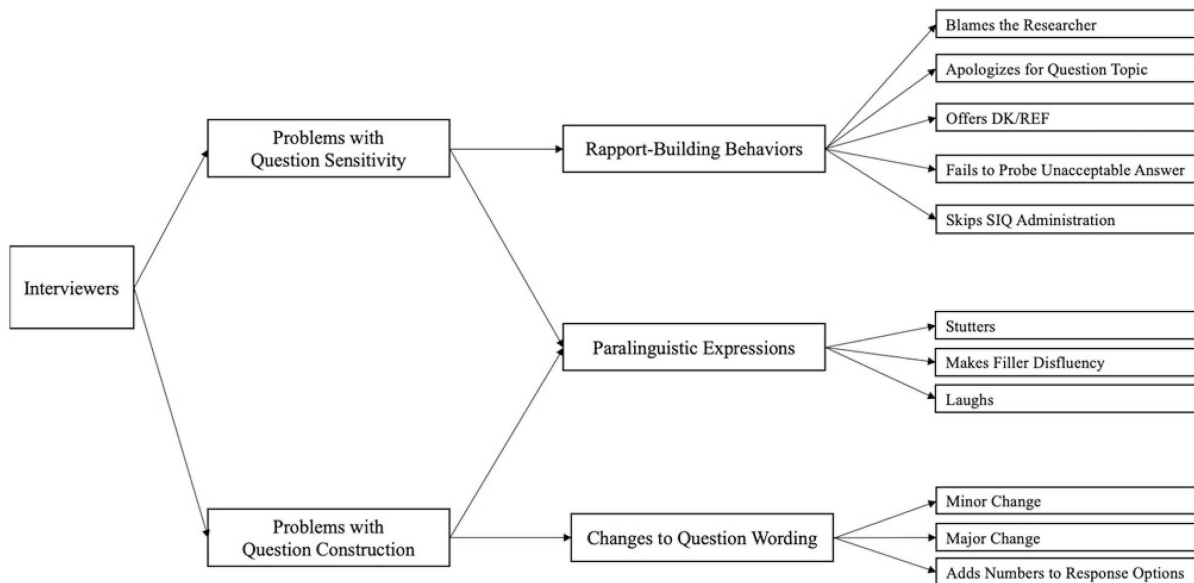


Fig. 3 Tree Diagram of Coded Interviewer Behaviors

Lord made [them] to be.” 1.2% of respondents made explicit comments about being uncomfortable with homosexuality (e.g., “[I’m] not involved in all this gay activities. I don’t even want to hear none about that.”). Overall, these backlash comments were made exclusively by heterosexual respondents (53.8% of the comments) and respondents who never provided an acceptable answer (46.2%). None of the respondents expressing backlash refused to answer the SIQ. Overall, 5.1% of respondents expressed at least one of our three indicators of question sensitivity (i.e., refusal to answer, intrusiveness, or backlash).

We find that interviewers express some indication of sensitivity (i.e., *rapport-building* behaviors) in 5.3% of the cases. These behaviors are displayed graphically in the top path of **Fig. 3**. In a few cases (0.7%) interviewers blame the researcher for having to ask the question (e.g., “They make us. I have to hear you say it.”). In only 0.3% of cases, interviewers apologized for the question topic. Next, in 0.4% of cases, interviewers offered the DK/REF response option to the respondent (e.g., “All you have to do is say skip it, I can move on.”).

The most common indicator of interviewers’ sensitivity to the SIQ was a failure to probe unacceptable answers, occurring in 40 cases (4.5%). For example, after a respondent answered “I’m normal,” the

interviewer replied with “What uh, what, what uh, your, your uh, uh [laughter] uh never mind let me just skip this one here,” and moved on to the next question. This finding implies that for these 40 cases that were not properly probed, interviewers should not have been able to code respondents’ answers, but still did. Of these 40 cases, interviewers classified 20 respondents as heterosexual, 1 as bisexual, 3 as “Other”, and 16 as DK/REF. As mentioned in the Data AND Methods section, we also find that interviewers failed to administer the SIQ in four cases (not included in our main analysis). In each of these four cases, the respondent expressed backlash in the question preceding the SIQ, which asked the respondent about the sex of their spouse or partner.

Indicators of Question Construction

As shown in Table 1, in 9.8% of our cases, respondents exhibited at least one indicator of difficulty with the construction of the SIQ (bottom path of Fig. 2). These expressions generally came from heterosexual respondents (67.1% of the comments) or respondents who never provided an acceptable answer (29.6%), though a small proportion came from respondents who refused to answer the SIQ (2.2%) or bisexual respondents (1.1%). Respondents self-identifying as gay or lesbian did not express any trouble with the SIQ’s construction.

Few respondents (0.2%) expressed concern that they had answered this question in the previous item about marriage. This indicates that respondents generally had no trouble with the placement of the SIQ after a question asking about relationship status and the gender of their partner. When examining respondents’ *confusion with the definition* of words in the SIQ, we find that 2.1% of respondents expressed confusion about the difference between heterosexual and straight. Few respondents (0.6%) confused sexual identity with sexual behavior. For example, these individuals sometimes expressed a lack of sexual desire at older ages (e.g., “At this point, I’m non-sexual at my ... old age.”). Other respondents (0.8%) confused sexual identity and gender (e.g., “I’m all woman, sweetheart.”), and 1.2% of respondents expressed confusion about the definition of any word in the question (e.g., “There’s so many different meanings on anything”). We also

find that several respondents were *confused about the appropriate response option*. 1.9% of respondents answered with a positive answer (e.g., “yes”) instead of a response option, and 4.2% of respondents answered this question with a negation (e.g., “no”).

Overall, indicators of respondent problems with question construction ($M = 9.8\%$) occur 4.7 percentage points more often than indicators of problems with question sensitivity ($M = 5.1\%$), a statistically significant difference ($z = -3.77$; $p < .001$). This indicates that although some respondents do find the SIQ sensitive, more deviations from paradigmatic interactions are caused by the SIQ’s construction.

Finally, when examining interviewers’ *question asking* behaviors (bottom path of Fig. 3), we find that interviewers make minor changes to question wording for 7.2% of respondents, and make major changes for 3.9% of respondents. Additionally, for 4.8% of respondents (43 cases), interviewers added numbers to each of the response options in the question stem. When interviewers added these numbers, respondents always gave an acceptable answer, indicating that this addition may increase respondents’ understanding of the SIQ. Of the 43 respondents who received numbers with the SIQ’s response options, 35 (81.4%) answered using the words of the response option (e.g., “heterosexual or straight”); all 35 of these respondents self-identified as heterosexual. The remaining eight respondents (18.6%) answered using the number of the response option (e.g., “I’m one”). One of these eight respondents self-identified as a sexual minority, and the rest indicated that they were heterosexual.

Similar to respondents, interviewers’ indicators of problems with question construction ($M = 11.8\%$) occur 6.5 percentage points more often than indicators of problems with question sensitivity ($M = 5.3\%$), a statistically significant difference ($z = -4.99$; $p < .001$). We see that interviewers’ interaction problems on the SIQ caused by question wording outnumber those caused by sensitivity. We do note, however, that these interviewer question construction behaviors (e.g., adding numbers to the SIQ’s response options) may have improved data quality by helping respondents to provide accurate, codable answers. Interviewer behaviors indicating problems with question sensitivity (e.g., failure to probe unacceptable answers), while less prevalent, are more likely to harm data quality.

Indicators of Question Sensitivity or Construction

Several of our indicators do not uniquely signal problems with the SIQ's sensitivity or construction. First, 2.9% of respondents indicated that they were uncertain of their answer. It is unclear if these expressions of uncertainty are due to confusion with question construction or because respondents are uncertain of their own sexuality. For example, many respondents indicated that their answer was "probably heterosexual" or "straight, I guess." One interpretation of these answers is that respondents are sure of their own sexual identity, but they are unsure that they have mapped that identity onto the correct response option. Another interpretation is that respondents are unsure of their own sexuality. Hedging phrases such as "probably" or "I guess" may be expressing some degree of uncertainty or sexual fluidity not being captured by the question's response options.

Among the indicators of dis-identification, we find that a small percentage of respondents (0.9%) answered the question in terms of what they or not (e.g., "I'm not gay."). The most prominent indicator of dis-identification was respondents expressing that they were fully straight (4.2%). Respondents used phrases that ranged from "I'm definitely straight" to "Strictly dickly here." 1.5% of respondents answered using heteronormative language, with most saying "I'm normal" to indicate straight.

Respondents' and interviewers' paralinguistic expressions may indicate either discomfort with the topic of the SIQ, or trouble understanding and responding to the question. Overall, 5.0% of respondents stuttered, 15.2% made filler disfluencies such as "uh" or "um", and 12.5% laughed. Interviewers stuttered for 3.5% of respondents, made filler disfluencies for 12.7% of respondents, and laughed for 9.1% of respondents.

Overall, we find that in 31.6% of cases, the respondent or the interviewer expressed some behavior indicating problems with the SIQ's sensitivity or construction (excluding paralinguistic expressions). This summary statistic demonstrates that when aggregating our individual indicators, sensitivity and construction issues with the SIQ may affect measurement of sexual identity for nearly one-third of respondents.

Associations Between Respondent and Interviewer Characteristics and Problems with the SIQ

We begin our bivariate analyses by examining the relationship between respondent age⁵ and indicators of sensitivity and construction problems with the SIQ (**Table 3**). We find a significant linear effect for three of our indicator groups: older respondents are more likely to express indicators of sensitivity ($t = 2.69, p < .05$) and construction problems ($t = 3.94, p < .01$), and have interviewer indicators of sensitivity ($t = 3.04, p < .01$). This indicates that, as hypothesized, older respondents are less likely to be comfortable answering questions about their sexuality and less likely to be familiar with the terms used in an SIQ. Additionally, interviewers likely react to older respondents' sensitivity by engaging in additional rapport-building behaviors. Respondent age is not significantly associated with interviewers' reactions to the SIQ's construction.

Next, interviewers are more likely to exhibit question sensitivity behaviors when respondents are less educated ($F = 4.96, p < .05$): these interviewer behaviors occur for 7.5% of respondents with a high school education or less, 6.4% of respondents with some college, but for only 2.9% of respondents with at least a college degree. Respondent indicators of sensitivity were not significantly associated with respondent education. These results indicate that interviewers are more likely to attempt to salvage rapport with less educated respondents when administering the SIQ, even though respondents' sensitivity behavior is not associated with their education levels. We also find that respondents with more education are less likely to have trouble with question construction ($F = 5.12, p < .05$). 13.6% of respondents with a high school education or less express issues with construction, compared to 9.0% of respondents with some college, and 7.6% of respondents with at least a college degree. Interestingly, interviewer construction-related behaviors were not associated with respondent education.

When examining interviewer characteristics (second half of Table 3), we find that experienced interviewers exhibit more problems with the SIQ's construction (e.g., changes to question wording) (26.7%)

⁵ Questions asking respondents about their age and education were administered before the SIQ in our questionnaire. Thus, interviewers knew these respondent characteristics when asking the SIQ.

Table 3 Indicators of Sensitivity and Construction Problems by Respondent and Interviewer Characteristics

	<i>Sensitivity: R</i>			<i>Sensitivity: I'wer</i>			<i>Construction: R</i>			<i>Construction: I'wer</i>		
	Mean	(SE)	F/t	Mean	(SE)	F/t	Mean	(SE)	F/t	Mean	(SE)	F/t
R characteristics												
R age												
18–35	1.2%	(0.9%)		1.8%	(1.0%)		3.0%	(1.2%)		15.0%	(4.2%)	
36–65	6.4%	(0.9%)		4.9%	(1.0%)		8.9%	(1.6%)		11.5%	(3.7%)	
66+	5.4%	(1.4%)	2.55	8.1%	(2.0%)	2.81	15.9%	(2.8%)	6.68**	10.5%	(4.0%)	1.61
Linear effect			2.69*			3.04**			3.94**			–1.22
R education												
HS or less	4.3%	(1.1%)		7.5%	(1.4%)		13.6%	(1.8%)		12.2%	(3.7%)	
Some college	3.4%	(1.4%)		6.4%	(1.7%)		9.0%	(1.9%)		9.8%	(3.5%)	
College+	6.8%	(1.1%)	1.81	2.9%	(0.9%)	4.96*	7.6%	(1.3%)	5.12*	12.9%	(4.5%)	0.36
Interviewer characteristics												
Interviewer tenure												
< = 1 year	5.6%	(0.6%)		4.5%	(0.8%)		10.0%	(1.1%)		6.6%	(1.7%)	
> 1 year	3.9%	(1.1%)	1.39	7.3%	(2.0%)	2.08	9.5%	(2.1%)	0.04	26.7%	(11.3%)	6.58*
Interviewer sex												
Male	5.4%	(0.8%)		6.0%	(1.3%)		10.9%	(1.5%)		12.1%	(5.9%)	
Female	4.7%	(0.8%)	0.40	4.2%	(0.8%)	1.58	8.4%	(1.3%)	–1.28	11.5%	(3.3%)	–0.08
Within-survey experience												
First half	4.1%	(0.8%)		4.3%	(1.0%)		9.5%	(1.4%)		10.9%	(3.6%)	
Second half	6.2%	(0.9%)	1.67	6.2%	(1.0%)	1.61	10.1%	(1.1%)	0.40	12.8%	(4.4%)	0.60

* $p < .05$; ** $p < .01$; *** $p < .001$

compared to inexperienced interviewers (6.6%; $F = 6.58$, $p < .05$). This difference indicates that more experienced interviewers may be more likely to repair the wording of the SIQ to make it more understandable. We also find that interviewer sex and within-survey experience⁶ were not significantly related to respondent or interviewer expressions regarding the SIQ's sensitivity or construction.

Discussion

In this study, we perform interaction-based behavior coding on a telephone administration of a question asking about respondents' sexual identity. This is the first study to our knowledge that explores the

⁶ These results change in a multivariate context. Using a logistic regression model to control for all other respondent and interviewer characteristics, we find that interviewers administering the second half of their WLT2 cases are more likely to engage in rapport-building, sensitivity behaviors than those administering the first half (coef. = .527, $p < .05$).

interaction between respondents and interviewers on an SIQ in a field study. We have four main takeaway points.

First, we find that some respondents conveyed concerns about the sensitivity of the SIQ. These concerns generally took the form of comments about the intrusive/ sensitive nature of the question. Such behaviors may signal respondents' uneasiness with being asked about a personal topic, but these comments might also be driven by an underlying discomfort with homosexuality and bisexuality (i.e., a form of backlash). Further research is needed to disentangle the possible motivations behind these comments about intrusiveness. Other respondents explicitly expressed anti-LGB backlash in the form of comments about their discomfort with homosexuality. Overall, 1.8% of our respondents refused to answer the question, a rate comparable to Ridolfo et al. (2012). These indicators of problems with the SIQ's sensitivity generally came from older respondents; thus, these behaviors may reflect generational differences in attitudes towards LGB individuals that might disappear over time.

Second, we find behaviors indicating that some interviewers also experience the SIQ as sensitive. This sensitivity was rarely expressed in the form of explicit comments about the SIQ (e.g., apologizing to the respondent for having to administer the SIQ). Rather, interviewers engaged in actions that limited the amount of time they spent interacting with the respondent on the SIQ. In 4.5% of cases, interviewers did not probe unacceptable respondent answers, but still coded final answers (i.e., interviewers potentially recorded incorrect sexual identities in these cases). A further four survey respondents were omitted from our analysis altogether because the interviewer skipped administration of the SIQ, but still coded a final answer for these respondents. Although these interviewer behaviors occurred in a relatively small number of cases, sexual minority individuals represent a small proportion of the population. Thus, interviewers intentionally misclassifying respondents in these ways may have serious implications for sample estimates of LGB respondents in telephone surveys.

We recommend that survey administrators better train interviewers to administer SIQs. Specifically, these results provide a collection of respondent behaviors that interviewers can expect to encounter when administering an SIQ. Interviewers can be trained to properly react to each of these behaviors in a way that continues the survey

interview. For example, when respondents comment that a question is off-topic or sensitive, or refuse to answer the question, interviewers can inform respondents that the question is for demographic purposes, and reinforce the survey agency's promise of privacy and/or confidentiality. Armed with this knowledge, interviewers may be less likely to avoid interacting with respondents on this question. Future work is needed to test whether such a strategy improves measurement with a telephone-administered SIQ.

Third, we find that respondent and interviewer behaviors indicating problems with the SIQ's construction were more prevalent than behaviors indicating concern over sensitivity. For example, some respondents expressed confusion over definitions of the response options in the SIQ, especially the difference between heterosexual and straight. Thus, we provide field evidence in support of previous researchers' calls to remove "heterosexual" from the response options of SIQs (e.g., Miller and Ryan 2011).

We also find that respondents sometimes interpret the SIQ as a yes/no question. To combat this confusion, it seems that interviewers sometimes change the wording of the SIQ by speaking numeric labels before each response option. In our survey, this change ensured that all respondents gave an acceptable answer. Interviewers with more time on the job were more likely to change their reading of the SIQ from the wording in the questionnaire. Thus, these interviewers may be relying on their experience administering the SIQ to help respondents better understand the question.

Fourth and finally, we identify several indicators that do not uniquely identify problems with the SIQ's sensitivity or construction. Several respondents elaborate on their answer to the SIQ with some degree of uncertainty (e.g., "I'm straight, I guess.") Additionally, respondent indicators of dis-identification could signal that respondents are either distancing themselves from the LGB identity, or dis-identifying from the LGB identity because being "not gay" is more salient than their heterosexual identity. Future research on this topic should ask interviewers to probe these responses to help disentangle the substantive meaning of these comments.

Based on this collection of evidence and previous research (e.g., Miller & Ryan 2011; Ridolfo et al. 2011, 2012; Dahlhamer et al. 2014), we propose a new interviewer-administered SIQ for testing: "*Do you*

consider yourself to be one, lesbian or gay, two, straight, that is not lesbian or gay, three, bisexual, or four, something else?" We believe that this will reduce respondent confusion over differences between the response options, may prevent respondents from prematurely answering with unacceptable positive (e.g., "yes") or negative (e.g., "no") answers, and may also reduce unacceptable dis-identification answers (e.g., "I'm not gay"). However, future research is needed to examine the measurement and data quality implications of this new SIQ construction.

This study is not without limitations. We have no benchmark data against which our present findings might be compared. These results would be strengthened by analyzing similar variables for non-sensitive questions (e.g., do interviewers incorrectly code respondent answers to non-sensitive questions at the same rate as the SIQ?). Additionally, we did not have access to several respondent and interviewer characteristics that may have been associated with our variables of interest. Variables such as respondents' religion or political affiliation, or interviewers' age should be investigated in future studies to better understand why respondents and interviewers might be exhibiting these behaviors.

Overall, this study presents several insights into the administration of an SIQ that would not have been possible using traditional methods that analyze patterns of response (e.g., item nonresponse rates) or use cognitive interviews. We find that nearly one-third of our respondents experience problems with the sensitivity or construction of the SIQ, and these problems come from both the interviewers and the respondents themselves. Administrators of telephone surveys can use these results to prepare their interviewers for respondent concerns about the SIQ's sensitivity. Questionnaire designers can use respondent and interviewer behaviors regarding question construction to revise the SIQ. Application of these findings will ensure that collection of sexual identity information via telephone surveys is not only feasible, but efficient and effective.

References

- Anderson, E. (2009). *Inclusive masculinity: The changing nature of masculinities*. Oxford, UK: Routledge.
- Adamczyk, A., & Pitt, C. (2009). Shaping attitudes about homosexuality: The role of religion and cultural context. *Social Science Research*, 38(2), 338–351.
- Black, D. A., Sanders, S. G., & Taylor, L. J. (2007). The economics of lesbian and gay families. *Journal of Economic Perspectives*, 21(2), 53–70.
- Boehmer, U., Bowen, D. J., & Bauer, G. R. (2007). Overweight and obesity in sexual-minority women: Evidence from population-based data. *American Journal of Public Health*, 97(6), 1134–1140.
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, 12, 389–437.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27, 703.
- Conron, K. J., Mimiaga, M. J., & Landers, S. J. (2008). *A health profile of Massachusetts adults by sexual orientation identity: Results from the 2001–2006 Behavioral Risk Factor Surveillance System surveys*. Commonwealth of Massachusetts, Dept. of Public Health.
- Dahlhamer, J. M., Galinsky, A. M., Joestl, S. S. & Ward, B. W. (2014). Sexual orientation in the 2013 national health interview survey: A quality assessment. *Vital and Health Statistics*, 2(169), 1–32.
- de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21(2), 233–255.
- Diamond, L. M. (2008). *Sexual fluidity*. Cambridge: Harvard University Press.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. Hoboken: Wiley.
- Dijkstra, W. (1987). Interviewing style and respondent behavior: An experimental study of the survey-interview. *Sociological Methods & Research*, 16(2), 309–334.
- Dykema, J., Schaeffer, N. C., Garbarski, D., & Hout, M. (2020). The role of question characteristics in designing and evaluating survey questions. In P. Beatty et al. (Eds.), *Advances in questionnaire design, development, evaluation, and testing* (pp. 117–152). Hoboken: Wiley.
- Eliason, M. J., Radix, A., McElroy, J. A., Garbers, S., & Haynes, S. G. (2016). The ‘something else’ of sexual orientation: Measuring sexual identities of older lesbian and bisexual women using National Health Interview survey questions. *Women’s Health Issues*, 26, S71–S80.
- Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 15–36). San Francisco: Jossey-Bass.

- Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. Sage.
- Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Thousand Oaks: Sage.
- Fredriksen-Goldsen, K. I., & Kim, H.-J. (2015). Count me in: Response to sexual orientation measures among older adults. *Research on Aging*, 37(5), 464–480.
- Gagnon, J. H., & Simon, W. (1973). *Sexual conduct: The social origins of human sexuality*. Chicago: Aldine.
- Gamson, J. (1995). Must identity movements self-destruct? A queer dilemma. *Social Problems*, 42 (3), 390–407.
- Garbarski, D., Schaeffer, N. C., & Dykema, J. (2016). Interviewing practices, conversational practices, and rapport: Responsiveness and engagement in the standardized survey interview. *Sociological Methodology*, 46(1), 1–38.
- Gates, G. J. (2017). *In U.S., more adults identifying as LGBT*. Available from <http://news.gallup.com/poll/201731/lgbt-identification-rises.aspx>
- Goffman, E. (1959). *The presentation of self in everyday life*. New York: Doubleday Anchor Books.
- Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. New York: Simon and Schuster.
- Haan, M., Ongena, Y., & Huiskes, M. (2013). Interviewers' question rewording: Not always a bad thing. In P. Winker, N. Menold, R. Porst, & D. Peter Lang (Eds.), *Interviewers' deviations in surveys* (pp. 173–193).
- Hatzenbuehler, M. L. (2014). Structural stigma and the health of lesbian, gay, and bisexual populations. *Current Directions in Psychological Science*, 23(2), 127–132.
- Hatzenbuehler, M. L. (2017). The influence of state laws on the mental health of sexual minority youth. *JAMA Pediatrics*, 171(4), 322–324.
- Herek, G. M. (2002). Gender gaps in public opinion about lesbians and gay men. *Public Opinion Quarterly*, 66(1), 40–66.
- Herek, G. M. (2009). Hate crimes and stigma-related experiences among sexual minority adults in the United States: Prevalence estimates from a national probability sample. *Journal of Interpersonal Violence*, 24(1), 54–74.
- Herek, G. M. (2011). Anti-equality marriage amendments and sexual stigma. *Journal of Social Issues*, 67(2), 413–426.
- Hooten, M. A., Noeva, K., & Hammonds, F. (2009). The effects of homosexual imagery in advertisements on brand perception and purchase intention. *Social Behavior and Personality: An International Journal*, 37(9), 1231–1238.
- Houtkoop-Steenstra, H. (2000). *Interaction and the standardized survey interview: The living questionnaire*. Cambridge: Cambridge University Press.
- Hox, J. J. (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological Methods & Research*, 22(3), 300–318.
- Japac, L. (2008). Interviewer error and interviewer burden. In J. M. Lepkowski et al. (Eds.), *Advances in telephone survey methodology* (pp. 187–211). Hoboken: Wiley.

- Kemph, B. T., & Kasser, T. (1996). Effects of sexual orientation of interviewer on expressed attitudes toward male homosexuality. *The Journal of Social Psychology, 136*(3), 401-403.
- Kosciw, J. G., Palmer, N. A., & Kull, R. M. (2015). Reflecting resiliency: Openness about sexual orientation and/or gender identity and its relationship to well-being and educational outcomes for LGBT students. *American Journal of Community Psychology, 55*(1-2), 167-178.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.
- Kulick, D. (2000). Gay and lesbian language. *Annual Review of Anthropology, 29*(1), 243-285.
- Lavin, D., & Maynard, D. W. (2001). Standardization vs. rapport: Respondent laughter and interviewer reaction during telephone surveys. *American Sociological Review, 66*(3), 453.
- Lavin, D., & Maynard, D. W. (2002). Standardization vs. rapport: How interviewers handle the laughter of respondents during telephone surveys. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 335-364). New York: Wiley.
- McLelland, M. (2000). Is there a Japanese 'Gay Identity'? *Culture, Health & Sexuality, 2*(4), 459-472.
- Michaels, S., et al. (2017). Improving measures of sexual and gender identity in English and Spanish to identify LGBT older adults in surveys. *LGBT Health, 4*(6), 412-418.
- Miller, K., & Ryan, J. M. (2011). *Design, development and testing of the NHIS sexual identity question*. Hyattsville, MD: National Center for Health Statistics.
- Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly, 75*(1), 99-114.
- Olson, K., & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly, 71*(2), 273-286.
- Olson, L. R., Cadge, W., & Harrison, J. T. (2006). Religion and public opinion about same-sex marriage. *Social Science Quarterly, 87*(2), 340-360.
- Perry, P. (2001). White means never having to say you're ethnic: White youth and the construction of "cultureless" identities. *Journal of Contemporary Ethnography, 30*(1), 56-91.
- Pew Research Center. (2017). *The partisan divide on political values grows even wider*. Available from <http://www.people-press.org/2017/10/05/5-homosexuality-gender-and-religion/>
- Powell, B., & Downey, D. B. (1997). Living in single-parent households: An investigation of the same-sex hypothesis. *American Sociological Review, 52*, 521-539.
- Ridolfo, H., & Schoua-Glusberg, A. (2009). Testing of NHANES A-CASI sexual behavior questions: Results of interviews conducted November 2008-February 2009. *Q-Bank*. Retrieved from <http://www.cdc.gov/qbank>

- Ridolfo, H., Miller, K., & Maitland, A. (2012). Measuring sexual identity using survey questionnaires: How valid are our measures? *Sexuality Research and Social Policy*, 9(2), 113–124.
- Ridolfo, H., Perez, K., & Miller, K. (2011). *Testing of sexual identity and health related questions, results of interviews conducted May–July 2005*. Hyattsville, MD: National Center for Health Statistics.
- Rotheram-Borus, M. J., et al. (1994). Sexual and substance use acts of gay and bisexual male adolescents in New York City. *Journal of Sex Research*, 31(1), 47–57.
- Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science*, 2(3), 179–183.
- Schaeffer, N. C., Dykema, J., & Maynard, D. W. (2010). Interviewers and interviewing. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 437–470). Bingley, UK: Academic.
- Schaeffer, N. C., & Maynard, D. W. (1996). From paradigm to prototype and back again: Interactive aspects of cognitive processing in standardized survey interviews. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 65–88). Jossey-Bass: San Francisco.
- SMART (2009). *Best practices for asking questions about sexual orientation on surveys*. Available from: <https://williamsinstitute.law.ucla.edu/wp-content/uploads/SMART-FINAL-Nov-2009.pdf>
- Stange, M. (2014). *Tailoring general population surveys to address participation and measurement challenges of surveying lesbian, gay, and bisexual people*. Doctoral dissertation, University of Nebraska-Lincoln.
- Stange, M., Smyth, J. D., & Olson, K. (2018). Drawing on LGB identity to encourage participation and disclosure of sexual orientation in surveys. *The Sociological Quarterly*, 1–21.
- Sylva, D., Rieger, G., Linsenmeier, J. A., & Bailey, J. M. (2010). Concealment of sexual orientation. *Archives of Sexual Behavior*, 39(1), 141–152.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859.
- Umberson, D., Thomeer, M. B., Kroeger, R. A., Lodge, A. C., & Xu, M. (2015). Challenges and opportunities for research on same-sex relationships. *Journal of Marriage and the Family*, 77 (1), 96–111.
- Valentine, J. (1997). Pots and pans: Identification of queer Japanese in terms of discrimination. In A. Livia & K. Hall (Eds.), *Queerly phrased: Language, gender, and sexuality* (pp. 95–114). Oxford: Oxford University Press.
- van der Zouwen, J., Dijkstra, W., & Smit, J. H. (1991). Studying respondent interviewer interaction: The relationship between interviewing style, interviewer behavior, and response behavior. In P. P. Biemer, R. M. Groves, L. E.

- Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 419–437). New York: Wiley.
- VanKim, N. A., Padilla, J. L., Lee, J. G., & Goldstein, A. O. (2010). Adding sexual orientation questions to statewide public health surveillance: New Mexico's experience. *American Journal of Public Health, 100*(12), 2392–2396.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*, 360–363.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks: Sage Publications.
- Yip, A. K. (2005). Queering religious texts: An exploration of British non-heterosexual Christians' and Muslims' strategy of constructing sexuality-affirming hermeneutics. *Sociology, 39*(1), 47–65.